AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE

General Editor
E.F.K. KOERNER
Zentrum für Allgemeine Sprachwissenschaft, Typologie
und Universalienforschung, Berlin
efk.koerner@rz.hu-berlin.de

Series IV – CURRENT ISSUES IN LINGUISTIC THEORY

Volume 311

Edited by Sergio Scalise and Irene Vogel (eds.)

Cross-Disciplinary Issues in Compounding

CROSS-DISCIPLINARY ISSUES IN COMPOUNDING

Edited by

SERGIO SCALISE
University of Bologna

IRENE VOGEL
University of Delaware

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

Synthetic compounds

With special reference to German*

Livio Gaeta
University of Naples "Federico II"

This chapter focuses on a traditional issue of word formation, namely synthetic compounds. The three basic approaches to the question will be reviewed and analyzed with the help of a large text corpus. The latter is of paramount importance for the analysis because dictionaries usually are not reliable for investigating highly productive word formation patterns and especially German compounding. It is shown that purely syntactic approaches do not cope well with the data, whereas a lexical approach like Construction Morphology is able to grasp the fine-grained distributional properties displayed by compounds. Furthermore, the corpus-based analysis allows us to shed some light on the complex network of semantic properties guiding the selective solidarity between deverbal head and nominal modifier by representing argument structure as a bundle of Dowty's base-roles.

1. Introduction

Synthetic compounds have a long history in the grammatical terminology, although the concepts of synthesis and of its cognate parasynthesis have been used with varying success in the different linguistic traditions. For instance, while in Italy (and partially in France), the concept of parasynthesis is fairly well established, the exact correspondents of the Italian parasynthetic derivatives are completely neglected in the English and German tradition – at least in modern times. On the other hand, our main object of investigation, namely synthetic compounds, are either completely neglected or simply assigned to compounding in Italy (and in France). To understand what is synthetic in a synthetic compound, it is usually referred to a double operation which

* This chapter results from a research developed within the PRIN-project COMPONET coordinated by Sergio Scalise (2005–07). I am deeply indebted with the editors and with an anonymous reviewer for helpful comments. Needless to say, opinions expressed and remaining mistakes are of my own responsability.

seems to take place at once: an operation of compound formation, in which the verb/noun relation naturally looks like being of an argumental nature, and an operation of deverbal noun formation. The latter may give rise to agent or action nouns, as in the German examples *Taxifahrer* 'taxi driver' and *Wasserladung* 'water loading'.

Actually, this interpretation of synthesis slightly diverges from the old grammatical usage, which basically translates into linguistic parlance the Kantian concept of synthetic judgements. The latter are distinct from the analythic ones because they imply the increase of the information content of the judgement, or assertion. Well aware of this usage, Tollemache (1945: 9) writes:

> "Se prendiamo a esaminare un composto greco per es. l'aggettivo *θεοφιλής*, ci accorgiamo che esso si compone di due temi forniti di desinenza aggettivale. Se proseguiamo il nostro studio, vedremo che da una parte è la desinenza che determina la natura grammaticale del nuovo composto, e che dell'altra questo è sintetico, contiene, cioè, più idee di quelle inerenti ai soli membri componenti … Ora, poiché per queste due proprietà esso si sottrae alle regole ordinarie della sintassi, i grammatici lo denominano asintattico".[1]

The part of the definition which is reminiscent of Kant's synthetic judgement refers to the fact that *θεοφιλής* 'dear to the gods' contains more ideas than those inherent in the single compound members. By this we can understand the argumental relation between head and modifier in the compound as in *Taxifahrer* above. On the other hand, the grammatical problem raised by *θεοφιλής* is given by the fact that the derivational operation causes the coming into existence of a new combination of morphemes, in that **φιλής* as such is not attested, whereas the adjective *φίλος* 'dear' occurs. This word formation mechanism is termed asyntactic because it cannot be simply explained away as resulting from the juxtaposition of morphemes, which is typical of a number of other compounds like Latin *respublica* 'republic', *terraemotus* 'earthquake', etc. (cf. Gaeta 2008). The latter sort of compounds is accordingly called syntactic and analytic. Thus, we have two different aspects of synthesis: on the one hand, a semantic "more", which is the argumental relation between the members, and on the other a formal "more", which is the special combination of morphemes, which does not occur outside of the synthetic formation. This latter aspect is shared by the so-called parasynthetic formations like It. *in-forn-are* 'to put into the oven'.[2]

1. [If we examine a Greek compound like the adj. *θεοφιλής*, we notice that it consists of two themes provided with an adjectival desinence. If we proceed with the analysis, we observe that on the one hand it is the desinence which determines the grammatical nature of the new compound, and on the other it is synthetic, namely, it contains more ideas than those inherent in the single compound members … Now, since these two properties keep this compound distinct from the normal rules of syntax, the grammarians call it asyntactic] (my translation).

2. Notice that since in the French tradition (cf. Arnaud 2004) complex words containing prepositions or particles (i.e., prefixes) are assigned to compounds, it turns out that what in Italy is considered to be a parasynthetic derivative is assigned to (para-)synthetic compounding in France.

At least, three possible analyses have been suggested to account for the simultaneous double operation traditionally labeled synthesis:

- Incorporation, i.e. lexical derivation via suffixation of a verb;
- Lexical derivation and subsequent composition;
- Lexical derivation via suffixation of a word group.

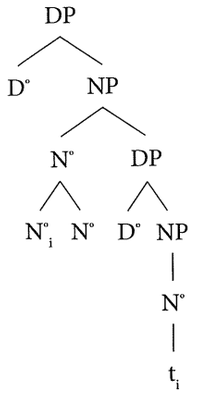
In what follows I will review the three different approaches with the aim of (i) establishing their adequacy to cope with synthesis and (ii) considering the theoretical implications which result from it. I will take German as main object of the investigation, because this language is traditionally known to be particularly rich in compounds. Furthermore, I could have access at large text corpora which will provide the empirical basis of my analysis. This is an essential aspect of the research, because corpora may help us overcome a general difficulty connected with the investigation of compounds and more in general with any extremely productive word formation pattern. If one does not want to rely solely on speakers' intuition and dictionaries are used, the problem arises that the latter usually discard the completely regular and transparent formations, because "dictionary-users need not check those words whose meaning is entirely predictable from its elements, which by definition is the case with productive formations" (Plag 1999: 96). In what follows, I will make use of a large text corpus (DEWAC01) developed within the Wacky-Project (cf. Bernardini, Baroni & Evert 2006) and containing about 170 million tokens extracted from the Web. Even though the reliability of the corpus is not optimal, in that it is difficult to evaluate a Web-corpus in terms of text types and distribution, its large size allows me to aspire to a certain significance of the results.

2. Incorporation in German?

Because of the enhanced polysynthetic potential exploited in the course of the German linguistic history (cf. Wurzel 1996), several proposals have been made for explaining synthetic compounds with the help of a mechanism of incorporation. This implies that the argumental relation between head and modifier is accounted for by making reference to an intermediate derivational stage, in which a verbal compound is generated. In fact, verbal compounds are usually considered to be the main characteristics of incorporating languages (cf. Aikhenvald 2007). There are two different versions of this approach, a radical and a moderate one.

The generalized incorporation suggested by Siebert (1999) translates the traditional idea of synthesis, which was asyntactic as shown above, into a strictly syntactic view. Morphology is equated with syntax: there are no (X-bar) principles specific for morphology. Accordingly, synthetic compounds are the result of syntactic head-movement, as depicted below (1999: 127):

(1) [*ein* [*Roman*$_i$*leser* [[t$_i$]$_{NP}$]$_{DP}$]$_{NP}$]$_{DP}$



The structural difference between a compound like *Romanleser* 'novel reader' and the corresponding phrase *der Leser des Romans* 'the reader of the novel' is merely given by the head-to-head movement of *Roman* to a Chomsky-adjoined higher node (the sister position of the DP). This extends to all compounds, independent of the deverbal nature of the head. Thus, a similar structural analysis is suggested for a compound like *Romanautor* 'novel author' with respect to the phrase *der Autor des Romans* 'the author of the novel', in which the head *Autor* cannot be claimed to be deverbal. This approach implies that the traditional difference between synthetic and analytic compounds is lost, and that every compound is taken to be synthetic. Moreover, incorporation is taken to occur independently of the true formation of a verbal compound.

Siebert tries to find support for her approach by making reference to general properties of incorporation which are supposed to be relevant for German compounds. Thus, Baker's (1988: 64) Government Transparency Corollary is called into question, whereby "[a] lexical category which has an item incorporated into it governs everything which the incorporated item governed in its original structural position". The effect of this corollary are taken to be mirrored in the compound *der Härte*$_i$*grad t*$_i$ *des Wassers* 'the degree of hardness of water' which corresponds to the phrase [*der Grad* [[*der Härte* [*des Wassers*]]]. However, this is not generally true, as shown by the impossibility of **Direktor*$_i$*foto t*$_i$ *der Schule* 'the photo of the director of the school'. To cope with this, Siebert's (1999: 142) tentative answer is to relate the effect of the corollary to the transparency of the head. Only abstract nouns are transparent enough to allow for argument inheritance, as shown by the following Google-examples:

(2) a. *In den Worten von Chinas Verhandlungs*$_i$*leiter t*$_i$ *mit der WTO, Long Yongtu,...*
'In the words of China's leader of the negotiation with the WTO, Long Yongtu,...'

b. *Was klingt wie eine Einladung für Piraten, sich Geiseln als unkomplizierte Einnahmequelle zu suchen, hat seinen Sinn im Krisenmanagement und der Verhandlungs*$_i$*leitung t*$_i$ *mit den Entführern.*
'What sounds as an invitation for pirats to look for hostages as an uncomplicated source of income, makes sense in crisis management and in the conduction of the negotiation with the kidnappers'.

Notice that this also implies that not every compound can be taken to be synthetic, but only those which are sensitive to argument structure, i.e. abstract nouns. In other words, a mechanism of lexical derivation must be postulated for at least a subset of compounds, although we are not precisely told which sort of compounds are lexically or syntactically derived.[3] In fact, while on the one hand a deadjectival noun like *Härte* is claimed to be sensitive to argumental inheritance, on the other a simplex like *Autor* must be taken to display an intrinsic argument structure, which is exploited at a syntactic level.

Finally, it can be objected against this strictly syntactic view that a governed adjective cannot be moved to a higher position, nor can it stay *in situ*, thus violating Baker's (1988) corollary:

(3) a. *Versuch eines tödlichen Attentats auf den Papst*
'attempt at an assassination of the pope'
b. **tödlicher*$_i$ *Attentats*$_j$*versuch t*$_i$ *t*$_j$ *auf den Papst*
c. **Attentats*$_i$*versuch tödlichen t*$_i$ *auf den Papst*

Notice that this restriction doesn't apply when the allegedly incorporated noun is a complex lexical unit like *generative Grammatiktheorie* 'generative grammar theory', which gives rise to a classical example of bracketing paradox (cf. Spencer 1991: 398–417). These paradoxes can only be explained by making reference to an autonomous lexical status of the phrase *generative Grammatik*, which is then combined into a compound (cf. Schlücker and Hüning 2009). In other words, *tödliches Attentat* is not lexical enough to give rise to such a bracketing paradox. Thus, the abstractness of the head cannot be the reason for incorporation to take place. Rather, the lexical status of the modifier is of relevance.

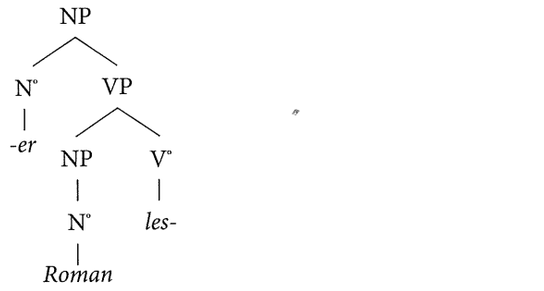
Besides the problems relating to the overgeneration induced by this approach, which also force us to assume a parallel mechanism of lexical compounding, there are other classical issues traditionally arisig from such strictly syntactic analyses, and in particular linking elements (*Fugenelemente*, cf. Fuhrhop 1996 among others). In this light, the general question for a strictly syntactic approach to compounds sounds: If

3. As pointed out by an anonymous reviewer, this is generally true for all compounds, because there are ambiguities between readings requiring the non-head as an argument and readings where this does not make sense. For instance, in a compound like *Alkoholfahrer* the non-argumental reading 'drunken driver' is clearly preferred over the argumental 'driver who transports alcohol', whereas in *Schmuckräuber* the argumental reading 'someone who robs jewellery' is preferred over the non-argumental 'thief wearing jewellery'.

compounding can be reduced to a purely syntactic operation, where are the allomorphic aspects of word formation to be dealt with? An answer can be sought in an approach sharing the separation hypothesis like split morphology, in which the actual derivation of compounds is conceived as separated from its "phonological" implementation (cf. Anderson 1992, Beard 1995). However, besides the standard arguments against split morphology (cf. Booij 1994), a specific problem is provided by German linking elements. On the one hand, there are cases which are productive and not relatable to the inflectional properties of lexical items like the so-called unparadigmic *-s-*: [[[V]-*ung*]*s*[N ~ Adj]$_Z$]$_Z$ (cf. *Bildungssystem* 'education system', *erklärungsbereit* 'ready to explain'). On the other, there are cases which appear to be strictly related to the inflectional properties of the modifiers, and also give rise to meaning effects: *Buchdecke* 'book cover' vs. *Bücherregal* 'books-shelf' (cf. Libben, Boniecki, Martha, Mittermann, Korecky-Kröll & Dressler 2009). It is not trivial to treat these alternations in a strictly split-morphology approach. At least, inherent vs. contextual inflection should be distinguished (cf. Booij 1996 and the discussion there).

Although similar objections of a formal nature apply to Rivet's (1999) moderate approach to incorporation, let us briefly review it, because it has at least the advantage of assuming incorporation only for deverbal nouns via an intermediate stage of verbal compounding:

(4) [$_N$ [$_V$ *Roman*$_N$ *les*$_V$] *-er*$_{NAf}$]



The crucial assumption in this approach is the empty intermediate step of a verb like **romanlesen*. Evidence in support of this analysis is claimed to come from other cases, in which the verbal compound is attested like *haushalten* 'to economize', *ehebrechen* 'to commit adultery', etc. (cf. Wurzel 1998 for a detailed investigation). Moreover, Rivet observes that in spite of the nonoccurrence of **romanlesen* as a verb, its nominalized form is attested, as in the following Google-example:

(5) *Hier führt Kant aus, daß das Romanlesen die Einheit des Denkens aufhebt.*
'Here Kant points out that reading novels destroys the unity of thought'.

This is surely true, but there is general consensus (cf. Wurzel 1998) that it is the nominal compound that provides the trigger for backderiving a verb, and not the other way

around. This explains fairly well the fact that only in a restricted number of cases a verbal compound is attested in the face of the large amount of nominal compounds headed by a deverbal noun or a nominalized infinitive. Thus, adopting Rivet's approach flows us again into the sea of overgeneration.[4]

3. Lexical derivation and subsequent compounding

A second way to deal with the two-faced nature of synthetic compounds is via derivation followed by compounding. Thus, a synthetic compound is taken to have the format [[N][V-Suff]$_Y$]$_Y$. As for the argumental relation between head and modifier, two different options are available: either lexical derivation is coupled with a syntax-driven operation which is responsible for argument inheritance, or argument inheritance simply results from semantics.

Oshita (1995: 180) adopts the first approach, and provides a lexico-syntactic definition of synthetic compound, which is "[a] compound whose nonhead satisfies the obligatory argument requirement of the head, irrespective of the latter's morphological origin". In his mind, English synthetic compounds like *whale-hunting* are accounted for by an alteration of the argument structure of the input verb, which is not suppressed by the morphological process:

(6) Suffixation of *-ing* (complex event)
a. Morphological Process:]$_V$ *-ing*]$_N$
b. A-Structure Alteration: (x (y)) → Ev (x = % (y))

Accordingly, "[t]he a-structure is retained, except that the original (external) argument x is now suppressed. The remaining argument y still needs to be projected" (Oshita 1995: 185). Notice that in this view only gerunds form synthetic compounds in English, whereas the cases like *taxi driver* are not synthetic, in that the head/modifier relation is not argumental. In the following representation "[b] expresses the nullification of the original a-structure with the angle brackets" (Oshita 1995: 186):

(7) Suffixation of *-er*
a. Morphological Process:]$_V$ *-er*]$_N$
b. A-Structure Alteration: (x (y)) → R = x ‹(x (y))›

Extending this view to German compounds, however, leads us into troubles, because both sorts of compounds manifest the possibility of preserving the subcategorization properties of the deverbal nouns, as we have seen above in (2).

Another problem of this approach is the role played by the so-called internal argument y, which – in Oshita's mind – basically is the direct object of transitive or the

4. There are other problems related to Rivet's approach, e.g. concerning an alleged condition on stress, but they will not be discussed here for the sake of brevity.

subject of unaccusative verbs. Accordingly, it should be automatically selected, provided that the argument structure of the verb displays it. As pointed out by Maling (2001: 455), however, this is not the case because of a number of transitive verbs which do not freely form synthetic compounds:[5]

(8) **Krankenbesucher* 'one who visits the sick'
**Königbediener* 'one who serves a king'
**Gasteinlader* 'one who invites a guest'
**Richterbitter* 'one who pleads the judge'

She argues that this is due to the semantic role associated with the direct object, which implies that the argument structure must be much richer than the rather poor frame based on syntactic functions suggested by Oshita. The semantic roles expressed in the argument structure determine the selectional restrictions of the synthetic compounds. In this light, internal arguments are not all equal: those linked with a goal role (i.e., in her view: recipients, experiencers and beneficiaries) encoded as accusative cannot form synthetic compounds.

In order to verify the suggestion that argument inheritance is sensitive to the semantic roles encoded by the syntactic functions, several verbs selecting a goal argument as a direct object were investigated in the corpus DEWAC01. The two verbs *einladen* 'to invite' and *besuchen* 'to visit' are representative respectively for a three-place predicate in which the direct object is linked with a goal and a further locative role is present, and for a two-place predicate in which the direct object is linked either with a goal or a locative:

(9) a. *Karl hat seinen Freund zu einer Party eingeladen.*
Karl has his-ACC friend to a-DAT party invited
'Karl has invited his friend to a party'.

b. *Karl hat seinen Freund besucht.*
Karl has his-ACC friend visited
'Karl has visited his friend'.

These findings are interesting from a double perspective. On the one hand, Maling's idea that semantic roles play a role in forming synthetic compounds seems to be confirmed, in that the selection clearly displays a preference for locatives instead of beneficiaries. On the other hand, this also requires to extend the idea that it's not argument selection that is relevant within a synthetic compound, but rather the semantics of the arguments involved. Thus, locatives are clearly preferred over beneficiaries, so that in the few divergent cases a locative interpretation is contextually available. This may give

5. This doesn't exclude that such compounds may be sporadically attested, as in the Google-example: *Der Krankenbesucher von der Personalabteilung hat heute versucht mit Ihnen zu sprechen* 'The visitor of the sicks from the personnel division has tried today to talk to you'. I thank an anonymous reviewer for this observation. As will be shown below, a corpus analysis is meant to partially solve the question of what is possible (and attested) and what is marginal or unattested.

Table 1. Synthetic compounds formed from *einladen* 'invite' and *besuchen* 'visit' in DEWAC01

[[X] *einlader*]	2	Locative: *Volksfest-* 'inviter to folk fests'	1
		Others: *Gefälligkeits-* 'pleasure i.'	1
[X [*einladung*]]	20	Agent: *Regierungs-* 'government's invitation'	1
		Beneficiary: *Selbst-* 'self i.', *Presse-* 'i. for the press'	2
		Locative: *Hochzeits-* 'i. to the wedding', *Konzert-* 'i. to a concert'	16
		Others: *Ferien-* 'i. during the holidays'	1
[N [*besucher*]]	166	Locative: *Konzert-* 'visitor of concerts', *Wien-* 'v. of Vienna'	157
		Others: *Durchschnitts-* 'average v.', *Zufalls-* 'occasional v.'	8
		Recipient: *Prostituierten-* 'v. of prostitutes'	1

rise to meaning effects like with *Prostituiertenbesucher* 'visitor of prostitutes', the only case of recipient in Table 1, in which prostitutes are degraded to a sort of locative!

Furthermore, notice that the preference for locatives over beneficiaries is independent of the syntactic function encoding the semantic role: it can be a direct object as in the case of *besuchen* or a prepositional object as in the case of *einladen*. This finding is at odds with most syntactically-oriented treatments of synthetic compounds, e.g. those based on incorporation, which usually allows only for structural cases to undergo head-to-head-movement of the kind depicted above, excluding the rest of the argument structure. Clearly, it is always possible to discard compounds like *Volksfesteinlader* 'inviter to public festivals' as generated via a different mechanism, which excludes the intervention of the reference to argument structure. This is also the line taken by Siebert and Oshita, as we have seen above. However, in the light of the argumental nature of the role encoded by the modifier, this move does not help much in assessing the status of these compounds, because it simply stipulates them to be different.

If semantics plays a crucial role in determining the selective possibilities of synthetic compounds, the question arises as for which sort of semantics is required. In this regard, Maling's view does not make predictions about the fact that except for beneficiaries the other semantic roles which are quite low in the thematic hierarchy are fairly well attested.

Instead of assuming a categorical restriction for goals as suggested by Maling, one may try a different approach to argument structure along the lines proposed by Primus

(1999).[6] She distinguishes three main proto-roles, which are hierarchically ordered, by making reference to Dowty's (1991) base-roles: Agent > Recipient > Patient. The latter reference is crucial for establishing such a hierarchy, in that a recipient is defined as the "first participant of a thematic predicate and this predicate is embedded in the thematic representation of the second participant of a higher thematic predicate such as CONTROL or CAUSE. The first property also characterizes Proto-Agents, the second property also specifies Proto-Patients" (Primus 1999: 55). This idea can be illustrated by the following representation which characterizes in broad terms the argument structure of two verbs displaying respectively recipients and beneficiaries:

(10) a. *Hans gab seiner Frau einen Apfel.*
Hans gave his-DAT wife a-ACC apple
'Hans gave his wife an apple'.
∀x∀y∀z[GIVE(x,y,z) → P-CONTROL(x, BECOME(POSS(y, z)))]

b. *Peter half seiner Frau.*
Peter helped his-DAT wife
'Peter helped his wife'.
∀x∀y[HELP(x,y) → P-CONTROL(x, BECOME(EXPER(y)))]

This approach allows us to look at semantic roles as bundles of interacting semantic properties or predicates. Accordingly, the semantic roles "are not distinguished by different basic thematic predicates, but only by their dependency relative to each other" (Primus 1999: 52). On this basis, we can formulate a preference to account for the tendency observed above for the selection of semantic roles which are quite low in the thematic hierarchy. Namely, the acceptability of synthetic compounds increases with the increase of the distance of the involved argument from the agentive prototype, which is defined as +human, +intentional, +individual, or, in Dowty's terms, which occurs as first argument of the highest base-roles.

From this preference, two predictions follow. The first prediction foresees that modifiers encoding semantic roles which are quite distant from the agentive prototype should be allowed in a synthetic compound. This can be shown to hold true on the basis of synthetic compounds containing verbs like *teilnehmen* 'to participate' and *fahnden* 'to search'. In the case of *teilnehmen*, the second argument encodes a locative role via a prepositional phrase, while the second argument of *fahnden* is a patient, encoded as a prepositional phrase:

(11) a. *Karl hat an der Tagung teilgenommen.*
Karl has at the-DAT conference participated
'Karl has participated at the conference'.

6. See Szigeti (2002) for an approach which also goes beyond mere argument inheritance in that it reflects the connection between word formation and the generation of concepts.

b. *Die Polizei hat nach dem gestohlenen Geld gefahndet.*
The police has after the-DAT stolen money searched
'The police has searched for the stolen money'.

As shown in Table 2, modifiers encoding semantic roles which are quite distant from the agentive proto-roles are in fact richly attested, irrespective of the syntactic function (direct or prepositional object, see also Table 1) present in the argument structure of the base verb.

The second prediction entails that dative-marked arguments should not be categorically excluded. Rather, they should be possible, displaying the same pattern observed so far of distance from a prototypical agent role. Recall that this is in contrast with what is claimed by Maling (2001), who sees a categorical restriction on the occurrence of the goal role in synthetic compounds. To assess this question, a detailed analysis of the synthetic compounds headed by the *-er*-derivatives of a typical dative-verb like *helfen* 'to help' proves useful (cf. (10b) above).

Although the majority of modifiers selects locatives (either in a concrete or in an abstract sense), a significant number of beneficiaries is attested. However, if we look at the single cases, we observe that they are distributed along well-defined semantic classes, which can be nicely captured by the idea laid down above of an agentive prototype characterized by a number of semantic properties. Recall that the preference entails that deviations are considered to be possible, but should accumulate at the margin of the prototype, thus displaying prototypical properties like + human, + intentional, + individual to a reduced degree. The beneficiaries occurring as modifiers

Table 2. Synthetic compounds formed from *teilnehmen* 'participate' and *fahnden* 'search' in DEWAC01

[[N] *teilnehmer*]	271	Locative: *Kongress-* 'congress participant', *Safari-* 'safari p.'	259
		Copulative: *Laien-* 'lay p.', *Pilot-* 'pilot p.'	4
		Others: *Dauer-* 'permanent p.', *Fach-* 'professional p.'	8
[X [*fahnder*]]	23	Patient: *Drogen-* 'drug searcher', *Kunst-* 'art s.'	19
		Others: *Ziel-* 'targeted s.', *Polizei-* 'police's s.'	4
[N [*fahndung*]]	33	Agent: *Interpol-* 'interpol search', *Polizei-* 'police s.'	2
		Patient: *Drogen-* 'drug s.', *Terroristen-* 'terrorists s.'	21
		Others: *Serien-* 'serial s.', *Sofort-* 'immediate s.'	10

Table 3. Synthetic compounds formed from *helfen* 'help' in DEWAC01

[[N] *helfer*]	184	Beneficiary	22
		Self compound: *Helfershelfer* 'accomplice'	
		Individuals: *Arzt-* 'doctor's assistant', *Pfarr-* 'pastor's a.', *Teufels-* 'devil's a.'	
		Groups: *Alten-* 'elderly a.', *Behinderten-* 'handicapped a.', *Flüchtlings-* 'refugee's a.', *Jugendgruppen-* 'youth a.', *Suchtkranken-* 'addict a.'	
		Collectives: *Gemeinde-* 'community a.', *Jugend-* 'youth a.', *Familien-* 'family a.', *Gestapo-*, *Luftwaffen-* 'air force a.', *Marine-* 'marine a.', *Nachbarschafts-* 'neighborhood a.', *Partei-* 'party a.', *Polizei-* 'police a.', *Rotkreuz-* 'Red Cross a.'	
		Abstracts: *Geist-* 'spirit helper'	
		Copulative	6
		Laien- 'lay assistant', *Bäcker-* 'a. baker', *Drucker-* 'a. printer', *Koch-* 'a. cook', *Maler-* 'a. painter', *Schreiner-* 'a. carpenter'	
		Abstract goal	107
		Locative	30
		Others	19

headed by *Helfer* are reported in Table 3. Apart from the self compound *Helfershelfer*, which is quite opaque, the modifiers are mainly given by groups (mostly formed by non-intentional individuals, like *Behinderte* or *Flüchtling*), or collectives. Notice that the only three cases of modifiers constituted by individuals are peculiar in that they might be partially treated on a par with what I tentatively classified as copulative compounds in Table 3 because their meaning comes close to be intersective rather than subordinative.[7] Thus, while an *Arzthelfer* cannot be categorized as a doctor, because it lacks some essential properties for being considered a doctor (for instance a completed degree), a *Kochhelfer* can be considered copulative if we perceive an assistant cook to be a cook, although still an apprentice. And, we can discuss whether a *Teufelshelfer* has undergone the process of becoming completely demonic or not.

Notice finally that most cases of modifiers headed by *Helfer* are abstract goals, like *Pflegehelfer* 'cure helper' or *Terrorhelfer* 'terror helper'. While this copes fairly well with the approach adopted here, because abstract goals are quite distant from the agentive prototype, this is problematic both for Maling's view and for the syntactically-based approaches which refers to incorporation. In fact, on the one hand an abstract goal should be categorically excluded anyway, unless it is considered a sort of locative, which is hard to claim in view of true cases of locatives like *Küchenhelfer* 'kitchen assistant'. On the other, such abstract roles are likely to encode true arguments, and not

7. See the discussion in Gaeta & Ricca (2009) about the continuum between the two different compound types.

merely adjuncts, as shown by the following Google-examples, in which the dative-marked argument is flanked by an instance of true adjunct:

(12) a. *Wird das der Pflege helfen [...]?*
'Will this help the cure?'

b. *Kinder können bei der Pflege helfen.*
'(lit.) Children can help at the cure'

This evidence supports the view of a semantic approach to synthetic compounds which is sensitive to the argument structure of the base verb. The argument structure must be understood in a wide sense, namely taking into consideration the semantic roles involved in the conceptualization of the event. The thematic hierarchy, irrespective of the syntactic coding, is a good predictor of synthetic compounds formation.

4. Lexical derivation via suffixation of a word group

Let us discuss the third possible approach to synthetic compounds, which is in a way the mirror-image of the incorporation approach depicted above, namely the lexical derivation via suffixation of a word group: $[[X\ Y]_Y\ Z]_Z$. Traditionally (cf. Erben 2006: 37–38), two types are distinguished in dependence of the relation between X and Y. This can be of a morphological or of a syntactic nature:

(13) a. *Zusammenbildung*: [[*Klavier-spiel*] *-er*] 'piano player'
b. *Zusammenrückung*: [[*saure Gurken*]*zeit*] 'time of sour gherkins'

In (13a) the synthetic compound results from the compounding of *Klavier* and *Spieler*. However, via a mechanism of backderivation a verb *klavierspielen* 'to play piano' is derived as hinted at above, which can then be reanalyzed as the base of *Klavierspieler*. The case of the so-called *Zusammenrückung* reminds us of the bracketing paradoxes mentioned above. The remarkable property here is that the adjective *sauer* 'sour' syntactically modifies the head *Zeit* (cf. *das Ende der sauren Gurkenzeit* 'the end of the time of sour gherkins', in which the adjective inflects in agreement with the head), but it clearly forms an onomasiological unit with the modifier *Gurke* 'gherkin' which must be stored as such in the lexicon.

We already pointed out above that backderivation has to be preferred over incorporation as a viable solution because the latter gives rise to severe problems of overgeneration, whereas on the other hand backderived verbs only occur in the presence of a nominal compound. The problem is that we also find cases in which a synthetic compound occurs which cannot be explained away as resulting from compounding, but needs to refer directly to a deverbal derivation. Thus, in *Todsagung* 'declaration of death' and *Verhaftnehmung* 'arrest' no deverbal nouns like **Sagung* or **Nehmung* occur based on *sagen* 'to say' and *nehmen* 'to take'.

Recently, research developed within the framework of Construction Morphology (Booij 2005, Gaeta 2006) has suggested to explain similar facts by making crucial reference to the idea of a rich lexicon, which contains not only idiomaticized expressions, but any kind of entrenched unit. On this basis, abstract schemas are analogically extracted, which may also be partially elaborated, and conflated with other schemas of a different nature. Thus, synthetic compounds like *Todsagung* result from the conflation of the schemas of compounding and of suffixation of abstract nouns:

(14) $[X\ X]_N$ $[[X]_V\ ung]_N$
\ /
$[X\ [[X]_V\ ung]_N]]_N$ 'the action of $[X\ X]_V$'
/ \
Todsagung *Verhaftnehmung*

As pointed out by Booij (2005: 129), "this template does not introduce a new formal type of complex words, but it expresses that it is the combination of two independently motivated word formation processes that systematically and productively co-occur".

If these premises are true, a number of predictions follow. First, we should expect that conflation effects should be sustained by recurrent input patterns, which favor the mechanism of entrenchment. As a matter of fact, the frequent activation of a certain of pattern enhances its direct access and sustains therefore its storage in the lexical memory. A second prediction we can make on the basis of a constructionist approach is that the single conflation patterns should be centered around highly entrenched models, which give rise to series of derivatives (*Reihenbildung*). Finally, because of their entrenchment, conflations should not strictly undergo principles like lexical blocking which have been shown to be highly sensitive to lexical frequency (cf. Rainer 1988).

As for the first prediction, we can search our corpus for cases like *Todsagung* above, and check for their type- and token-frequency. This is exemplified below on the basis of the conflation schema [X [V *ung*]] for the three highly frequent *geben* 'to give', verbs *machen* 'to make', *tun* 'to do', which do not normally form the respective abstract nouns **Gebung*, **Machung*, **Tuung*.

The entrenched pattern seems to be provided by the schema [X [*machung*]], which is robustly present with 127 different types. A detailed analysis of the token frequency

Table 4. Synthetic compounds with *-ung* formed from *machen* 'make', *tun* 'do' and *geben* 'give' in DEWAC01

	typ.		tok.
[[Adv] *machung*]	121		
[[N] *machung*]	6		
[[Adv] *tuung*]	1	*Genug-* 'enough'	331
[N [*gebung*]]	2	*Kund-* 'known'	1071
		Frei- 'free'	1

Table 5. Synthetic compounds with *-ung* formed from *machen* 'make', tun 'do' and *geben* 'give' and their basic recurrent patterns in DEWAC01

	tok.	$[Adv+V]_{V°}$	$[Adv\ V]_{V'}$
Bekanntmachung 'making known'	2814	756	987
Geltendmachung 'm. valid'	2156	1	8452
Kundmachung 'm. known'	907	8	20
Wiedergutmachung 'm. good again'	546	76	42
Glaubhaftmachung 'm. believable'	247	1	828
Mobilmachung 'm. mobile'	144	1	46
Rückgängigmachung 'm. regressive'	142	0	726
Zugänglichmachung 'm. accessible'	130	67	1249
Nutzbarmachung 'm. usable'	77	0	255
Sichtbarmachung 'm. visible'	58	0	468
Kenntlichmachung 'm. knowable'	54	2	206
Genugtuung 'satisfaction'	331	1	44
Kundgebung 'demonstration'	1071	25	5
Freigebung 'liberation'	1	1649	172

is presented in Table 5 above. For the other two cases, we record only two types, in which however respectively *Genugtuung* 'satisfaction' and *Kundgebung* 'demonstration' display a considerable token frequency. Let us see whether these synthetic compounds are likely to rely on highly recurrent patterns. Two possible input patterns occur, which are given below with the help of the usual Google-examples:

(15) a. $[Adv+V]_{V°}$ *bekanntmachen* 'to make known'
hat man sich miteinander bekanntgemacht.
'we introduced to each other'.

b. $[Adv\ V]_{V'}$ *bekannt machen* 'id.'
Filme, die Oliver Stone bekannt gemacht haben.
'Films, which have made Oliver Stone known'.

Clearly, the first pattern shows a higher lexical entrenchment than the second one. In Table 5 the single compounds are given with their respective token number and flanked by the token frequency of the two possible patterns on which they rely.[8]

8. For the sake of brevity, only the cases of the pattern [X [*machung*]] which display a token frequency higher than 50 are reported in the table. Moreover, only the adjacent sequences were searched, although German is known to display a highly constrained word order, in which the finite verb forms obligatorily occupy the second position in main declarative sentences. At any rate, including also these cases would only have the effect of increasing the frequency of the single patterns.

Table 6. Synthetic compounds formed from *geben* 'give' in DEWAC01

	typ.		typ.
[[N] *gebung*]	287	[N [*gabe*]]	113
[[Adj] *gebung*]	70	[Adj [*gabe*]]	80
Others	3	Others	26
Tot.	360	Tot.	219

On average, the frequency values seem to confirm the idea that the compounds are based on recurrent patterns. Furthermore, notice that conflation chain effects occur, as in the following cases:

(16) a. [[*Kampf*]*unfähig*[*machung*]] 'making unable to fight'
b. [*Wieder*[[*gut*]*machung*]] 'amend, lit. making good again'

Especially the second case is very telling of the idea of recurrent patterns because the intermediate derivational stage °*Gutmachung* is not attested in the corpus.

As for the second prediction, we may observe that conflation patterns usually crystallize starting from single models which are robustly attested and give rise to *Reihenbildungen*. By doing so, lexical blocking may be overcome because the recurrent pattern forces the instantiation of productive mechanisms of word formation.

Let us investigate in detail the synthetic compounds based on the verb *geben* 'to give'. In spite of the fact that the robust occurrence of the opaque apophonic derivative *Gabe* 'giving, gift' (1864 tokens) blocks the derivative *Gebung* 'id.', which nevertheless shows up twice in the corpus, the numbers in Table 6 are essentially different if the conflation patterns are taken into consideration.

The high number of compounds headed by *Gebung* is mainly given by conflation chains built on the most frequent model *Gesetzgebung*, which displays a high token frequency (2758) and gives rise to 221 compounds of the type [X [*gesetzgebung*], namely about 60% of the total. Notice that the opaque counterpart *Gesetzgabe* is not attested at all. Thus, the entrenched model gives rise to a considerable series of derivatives which in their turn reinforce the presence of the pattern in the lexicon. On the other hand, the compounds based on the opaque but frequent derivative *Gabe* are clearly recessive against the productivity of the conflation pattern.

5. Conclusion

To sum up, the first conclusion that comes to mind is that corpora are of growing relevance for morphological analysis. Especially for highly productive word formation patterns like German synthetic compounds, a corpus-based analysis may shed some light on the intricacies of their selectional properties and of their distribution. They

provide evidence for an approach based on the idea of a rich lexicon, which contains schemas of growing abstractness.

Second, this view is at odds with a strictly syntactic approach like incorporation, because the latter is not able to account for the fine-grained selectional properties of compounds. Instead, a framework provided with a rich semantic information like Dowty's base-roles is required for capturing the spectrum of possible modifiers within synthetic compounds.

Finally, corpora also provide invaluable data for assessing the degree of entrenchment of the single patterns which sustain the conflation of different word formation models which are well accomodated within the framework of Construction Morphology. This can be said to capture the essence of synthetic compounds like the Greek *θεοφιλής* seen at the start, which are at the heart of the lexicon-syntax interface.